

# Cohesion errors in the academic writing of second-language users of English

*Hilton Hubbard*  
*Department of Linguistics, Unisa*

It is not only English teachers, but rather all who regularly find themselves reading students' essays, reports, assignments and the like, who would agree heartily with Wallace Chafe (1986:12) when he says that "the number of individuals who ever learn to write well is impressively small".

How important is it that students should "write well"? Although it is very difficult to quantify the relationship between students' ability to write coherently and their levels of academic achievement, these two variables could be expected to correlate positively - an expectation strongly supported by one of the findings of a recent empirical study (Hubbard 1989). Given this correlation, and the present widespread concern about the first-year failure rate at South African universities, the need to improve students' writing skills becomes a matter of some urgency. The situation is exacerbated in this country by the rapidly growing number of students entering university for whom English, the language of the assignment and the examination paper, is not the mother tongue. To quote Widdowson (1983:34): "getting the better of words in writing is commonly a very hard struggle. And I am thinking now of words which are in one's own language. The struggle is all the greater when they are not."

## AIMS OF THE STUDY

We are unlikely to contribute much towards the improvement of students' writing by way of hit-or-miss, impressionistic remarks in the margins of their assignments. We need a more systematic approach to the teaching of academic writing skills, based on 'a better understanding of the linguistic features and rhetorical structures that create coherence as well as greater insight into the problems students experience in trying to use them' (Bamberg 1984: 305-306).

The study reported on here examines the use of one set of linguistic features that is of particular relevance to the business of text-making, i.e. the cohesion system of English, in the writing of students who are users of English as a second language. The primary aim of the study was to establish and apply a framework for the analysis of the cohesion errors made by the students, so as to throw more light on the main problem areas. A secondary aim was an essentially stylistic one: to compare the density of cohesion errors in assignments as opposed to examination answers, so as to establish whether the latter task-type does indeed tend to provoke more errors.

## METHOD

In the course of the year, all 23 assignments written by black students of the University of South Africa as answers to the following Linguistics I question were collected:

*Explain each of the nine situational variables identified by Hymes. Exemplify your discussion by referring to one of the following speech events: a courtroom trial; a conversation between friends; a school history lesson. The examples may be drawn from a language or languages of your choice.*

At the end of the year, 15 of the 23 students answered the comparable examination question:

*Explain how the following five of the nine variables identified by Hymes can affect the language used in a particular situation: setting; participants; key; instrumentality; and topic. Exemplify your discussion by referring to one of the following speech events: a conversation between friends; a school history lesson; a courtroom trial. In your answer you may refer to a language or languages of your choice.*

These 15 examination answers, together with the same 15 students' assignments (from the 23 originally collected) constituted the data corpus for this study. The first languages of the students were: Northern Sotho (6); Zulu (3); Tswana (3); Xhosa (2); and Southern Sotho (1).

The texts studied exemplify what might be called *student academic writing* - that sub-genre of expository writing which is required from students when writing about the content of their school, college or university subjects and which is therefore distinct from general composition writing, as practised in language departments. Student academic writing is, unlike "freshman composition", the kind of writing that is directly relevant to academic achievement and is thus also the focus of academic support programmes.

After the academic writing texts were selected, all cohesion errors were analysed. This meant that first the texts were segmented into their component textual units and then the cohesion errors were classified in terms of a formal taxonomy. The two most frequent formal categories of error - reference and conjunctive cohesion - were then also classified in terms of a functional taxonomy. Details of both types of classification are given in the next section.

## **ANALYTICAL FRAMEWORK**

The key elements of the framework applied in this study concern: (a) the English cohesion system; (b) the textual unit; and (c) the formal and functional error analysis taxonomies. Let us take each in turn.

## Cohesion

Halliday and Hasan's pioneering study, *Cohesion in English* (1976), helped to usher in a new interest in text linguistics, which has been one of the most prominent concerns of linguistics in the eighties. They define *cohesion* as what "occurs when the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it" (Halliday and Hasan 1976:4). Two elements that are related to one another in this way form a *cohesive tie*, and these ties can hold within sentences, as in

[1] *Evita sent her biltong to Bapetikosweti*

or they can straddle sentences, as in

[2] *Evita drove her Merc to Bapetikosweti. On arriving there, she waited for the slower police van to offload her biltong.*

In [1], the presupposing element or cohesive device, *her*, links back to *Evita* in the same sentence, while in the second sentence of [2] *she* and *her* link back to *Evita*, *there* links back to *Bapetikosweti*, and *slower* links back to *her Merc* in the previous sentence.

Halliday and Hasan's main concern is with cohesion across sentence boundaries, because these types of cohesive ties "are the ONLY source of texture, whereas within the sentence there are the structural relations as well" (1976:9). Practically all subsequent studies of cohesion have taken their cue from these authors and focused on the text-building role of cohesion as manifested between, rather than within, structural units. The present study is no exception in this respect, although - as will be explained shortly - the basic structural unit defined for cohesion analysis here is not the orthographic sentence, as is the case in Halliday and Hasan (1976).

Although cohesion is by no means an unproblematical notion, there is considerable agreement on Halliday and Hasan's five-way classification, i.e. reference, substitution, ellipsis, conjunction and lexical cohesion.

*Reference cohesion* occurs when certain types of items form ties with other items in such a way that "the same thing enters into the discourse a second time" (Halliday and Hasan 1976:31). Reference cohesion devices divide into three groups: pronominal, demonstrative and comparative reference. Pronominal reference cohesion devices are exemplified in [2] above by the items *she* and *her*, which share reference with *Evita*, while the demonstrative *there* shares reference with *Bapetikosweti*, and the comparative *slower* is interpreted relative to *her Merc*.

*Substitution cohesion* occurs where the presupposing item has the same meaning as the presupposed one(s), but not the same reference. Substitution can be nominal, as in

[3] *She got her leg of venison for nothing. I had to pay for my one.*

Or it can be verbal, as in

[4] *She cultivates friends in high places. Maybe that's what I'd better do too.*

Ellipsis cohesion is, in effect, "substitution by zero" (Halliday and Hasan 1976:142), as in

[5] *Have you been to Albert's game farm? I have*

where the meaning of the second sentence is interpreted in terms of the recoverable elements *been to Albert's game farm*.

*Conjunctive cohesion* occurs when certain types of items express semantic relation that specify "the way in which what is to follow is systematically connected to what has gone before" (Halliday and Hasan 1976:227). Formally speaking, conjunctives can be divided into:

coordinators (e.g. *and, but*);  
subordinators (e.g. *although, so that*); and  
adverbials (e.g. *however, for example*);

but in Halliday and Hasan (1976) priority is given to a semantic classification of conjunctives. Four basic classes are recognised:

additive (e.g. *and, for example*);  
adversative (e.g. *but, although, on the contrary*);  
causal (e.g. *so, therefore, as a result*); and  
temporal (e.g. *then, after that*).

*Lexical cohesion*, "the cohesive effect achieved by the selection of vocabulary" (Halliday and Hasan 1976:274), is usually the most frequently occurring form of cohesion in written text, but it is also the least adequately defined. All the same, the present study employed a modified version of Halliday and Hasan's subcategorisation of lexical cohesion:

repetition of the same item or appearance of a related word with the same root (e.g. *drive - driver - driven*);  
synonymy or near-synonymy (e.g. *graft - corruption*);  
hyponymy (e.g. *car - vehicle*) or "general noun";  
relationship (*car - thing*); and  
collocation (e.g. *car - drove - road*).

As can be seen in Table 1, the formal categorisation of cohesion errors was made in terms of the five major types of cohesion, and the main subtypes of each. Before turning to the error analysis, though, we need to consider - very briefly - the question of the textual unit defined in the framework of this study.

### Textual units

We saw earlier that although cohesion is possible within the sentence, Halliday and Hasan (1976) concentrate on intersentential cohesion because it is at this level that its role in text building is most important. How, though, should the basic textual building-

block be defined? There are problems attached to Halliday and Hasan's choice of the orthographic sentence as the basic textual unit. For example, this forces them, for purposes of analysis, to ignore *so* as an instance of conjunctive cohesion in [6], while accepting it in [7], despite the fact that the word plays an equivalent role in both:

- [6] *The policeman saw a suspicious-looking bakkie, so he stopped it.*
- [7] *The policeman saw a suspicious-looking bakkie. So he stopped it.*

Apart from the orthographic sentence, the most commonly used units in textual analysis have been the proposition, the T-unit (e.g. Hunt 1965) and the clause. In the present study a modified version of Lieber's (1981) "functional unit of discourse" or F-unit was found to be most satisfactory, particularly for analysing student writing. F-units include clauses and also certain phrasal structures that are functionally equivalent to clauses because they "serve identifiable rhetorical functions in the development of a discourse" (Lieber 1981:58). Thus in [8], for example, the phrase *but not too bad* represents a new turn in the discourse, and is considered to be a separate F-unit:

- [8] *It tastes gamy, but not too bad.*

Space constraints prevent further discussion of this unit (cf. Lieber 1981 for more detail, and Hubbard 1989 for modifications): the key point is, though, that for this study cohesion was defined in terms of ties holding across, and not within F-units, and cohesion errors were analysed accordingly. Let us look more closely at the framework used for the error analysis.

### **Error analysis of cohesion**

The analysis of any set of errors is a complicated enterprise involving problematic choices at all stages, i.e. when identifying, classifying and explaining the errors. Without doubt the stage that

is most fraught with difficulty and most resistant to objective analysis is error explanation (cf. Spolsky 1979 and Abbott 1980 for just some of the problems attached to this aspect of error analysis). In the absence of elicitation from the student writers of sufficiently rich data on their composing process, attempts at explaining errors purely on the basis of the written product are almost invariably simplistic. Given its focus on product rather than process, the purpose of this study was not to present systematic explanations of cohesion errors, but rather to identify problem areas by way of a useful classification of the errors.

In this study the classification of cohesion errors was approached from two different perspectives - one formal, one functional - which are reflected in Table 1 and Table 2 respectively.

In accordance with the first perspective, the errors were classified formally in terms of the various categories and subcategories of cohesive device discussed earlier.

Thus examples [9] and [10] are of nominal and verbal substitution errors, respectively:

- [9] *Many address forms are taboo. **One** that must not use them is the daughter-in-law.*
- [10] *In conversation between friends, the interruption does no harm, everyone is free to **do so**.*

Ellipsis errors were very rare. The same student responsible for [10] also produced

- [11] *If it was a lecture, questions could only be asked by the lecturer, the students couldn't.*

There were a number of lexical cohesion errors. In [12] we see the worst kind of elegant variation:

- [12] *In the school history lesson we have the following roles: pupil-pupil relations and pupil-teacher role.*



This error can be related to the lexical cohesion subcategory of repetition (see above), as repetition of *roles* or the use of other cohesion devices, such as substitution ellipsis, is called for. In

[13] ***Participants***

*This focuses on...*

a superordinate term to the hyponymous *participants*, such as *variable*, should have been included. Errors in a third subcategory of lexical cohesion, i.e. collocation, include

[14] *Because what does the teacher impart? The main thing is material to the pupils.*

*Impart* could collocate with content matter, but not with *material*.

Errors in the three main categories just discussed, i.e. substitution, ellipsis and lexical cohesion, were relatively infrequent in the texts studied. By far the most errors were found in the categories of reference and conjunctive cohesion. Largely for this reason, these errors formed the main focus of the study, and an attempt was made to supplement the formal classification of errors (as in Table 1) with a functional classification of the reference and conjunctive cohesion errors. This classification was functional in the sense that the errors were defined in terms of the different problems they cause for readers as they process the text, and these are given in Table 2. Let us now look more closely at the classification of the reference and conjunctive errors.

The reader-oriented functional classification distinguishes two main types of error. On the one hand, there are those errors which, although they make the reader's task more difficult, do not prevent him from eventually identifying a plausible interpretation for the incorrectly used item. On the other hand, there are those errors which leave the reader unable to identify a plausible interpretation, either because of an absence of such an interpretation, or - in the case of ambiguity - because there is more than one

possible interpretation. All told, six categories of error were posited (the strategy that the reader is presumed to have to employ in order to achieve an interpretation is given in parentheses):

Interpretation achievable:

- (a) extraction (derive interpretation by way of additional inferences);
- (b) form (reconstruct correct form in accordance with contextual clues);
- (c) omission (add cohesive item in accordance with contextual clues);
- (d) replacement (replace existing item with cohesive item in accordance with contextual clues).

Interpretation not achievable:

- (e) ambiguity;
- (f) no reference or relation.

Let us take each of these in turn.

### *Extraction*

This term was first used by Lieber, who defined it (1981:215) for reference cohesion in terms of items "for which a referent could be extracted, or derived, from a preceding or following phrase or longer segment of text. With extractions, a reader must provide an extra step, an extra segment that is not realized in the text...".

Although a more detailed subcategorisation of extraction has been developed for a separate study (Hubbard 1989), using Sanford and Garrod's (1981) account of text processing as the point of departure, in the present short report we will not be able to go into such detail. A reference cohesion error that is classifiable as an extraction is exemplified in:

- [15] *A general topic for a conversation between friends can be people of the opposite sex, but the specific underlying aim can be to get the views about the girl **he** seduced.*

Here, although processing is made unnecessarily difficult, a referent for *he* can be derived by the reader from knowledge of the situation sketched earlier in the text: one of the "friends" participating in the conversation is being referred to.

The notion of extraction can be defined for conjunctive as well as reference cohesion, i.e. by replacing "a referent" in Lieber's above definition with "an interpretation". Example [16] includes a conjunctive extraction error:

- [16] *A happy mood is necessary to make everybody at easy [sic] and thus with all this the language use will be formal and selected. We do not expect to find a teacher threatening or scolding during such a discourse.*

The semantic relation of reason-result (see Crombie 1985a; 1985b) normally signalled by *thus* initially seems improbable here because of conflict between the concepts 'happy mood' and 'being at ease' on the one hand, and 'formal language use' on the other. It is only on reading the second sentence that one realises that this contrast is being glossed over in favour of a contrast between 'formal language use' and 'threatening or scolding' which, it is implied, is some kind of non-formal language use. Once again, the reader has to make an inferential leap to "extract" the writer's meaning.

### *Form*

This category was postulated to cover all cases where there is some error in the form of the cohesion item itself. In the domain of reference cohesion, the most common errors of this sort involved number or agreement problems, as in

- [17] *The term "ends" relates to the aims and purposes that characterise the speech event. **This** could be to inform or to be friendly.*

Occasionally, conjunctives were incorrectly expressed, as in

[18] ...*a humorous person, for an example.*

### *Omission*

This type of error is analysed when a cohesion item is omitted in an obligatory environment, as in

[19] *The category through which communication takes place is the channel. In a short history lesson the spoken medium can be associated with face-to-face channel.*

where the missing reference item is *the*, and in

[20] *These are: that one must not interrupt, that normal voice should not be used except when scheduled in a choir singing.*

where the missing conjunctive item is *and* (a very common error).

### *Replacement*

This type of error involves substitution of an inappropriate item for the required one, and is something of a "wastebasket" category, in that it is applied to all cohesion errors where an interpretation can be achieved by the reader but which cannot be classified as extraction, form or omission errors (It is true that there can be an element of "replacement" in the latter three error types as well, but each of these cases have further defining characteristics).

An example of a reference replacement error is

[21] *A courtroom trial is a formal situation. We see it in the roles of the participants.*

where *this* rather than *it* is required.

The most common of the conjunctive replacement errors involved the use of an item which could function as a conjunctive, but was not the one intended, as in

[22] *He also write **that** they can read.*

where it is clear from the context that *so that* is intended ([22] does of course represent an archaic use of *that* - quite possibly the student has been influenced by the Authorised Version of the Bible).

### *Ambiguity*

This is the first of the two categories of cohesion error which prevent the reader from arriving at a plausible interpretation, and as such it is reasonable to suppose that these categories will tend to have a more serious effect on coherence than the four types of error discussed above.

There were no examples in the corpus of conjunctive ambiguity, but one of the reference ambiguities is exemplified in

[23] *The language of the judge and the defendant is not the same. **He** will be more polite.*

### *No interpretation*

An example of complete absence of a plausible referent for a reference item is

[24] *In **other** languages the role relationship of participants is most obviously reflected in their choice of second person pronouns and address forms.*

Although the comparative reference item *other* was used, there was in this text no previous mention of a language or even of language data.

The use of conjunctives to signal non-existent relations between textual units, as in [25], was fairly common:

[25] *Topic as one of the nine situational variables refers to the topic discussed in a history lesson. There are **therefore** general and specific topics.*

Let us now consider the results of the application of the analytical framework to the corpus of student assignments and examination answers studied.

## FINDINGS

The results of the analysis are presented in Tables 1 and 2.

The total corpus comprised approximately 16 500 words, of which the assignment texts accounted for about 11 000 and the examination texts for about 5 500 words. The figures in the "% Density" columns of Table 1 are arrived at by dividing the number of errors made by the number of words in the relevant set of texts. The resulting figure indicates what percentage of the words used were cohesion errors.

One of the salient features of Table 1 is the consistently higher frequency of errors in the examination texts, which are more important to academic success than assignments. Thus for every 100 words of examination text an average of 1,49 cohesion errors were committed, while only 0,85 such errors appeared over the same length of assignment text. There are of course a number of plausible reasons for this, including the unavailability of the source texts and the extra time pressures in the examination, which would inhibit language monitoring (see e.g. Krashen 1982).

Of the five main categories of cohesion, only ellipsis revealed a higher density of errors in the assignment texts, but the numbers here were so small (2 errors, as against none in the examination texts) as to be of no statistical significance whatsoever. Lexical errors, on the other hand, were most noticeably more frequent in

TABLE 1: FREQUENCIES AND % DENSITIES OF ERRORS (BY COHESIVE DEVICE)

|               | FREQUENCY |      |       | % DENSITY |      |       |
|---------------|-----------|------|-------|-----------|------|-------|
|               | Ass.      | Exam | Total | Ass.      | Exam | Total |
| ALL ERRORS    | 95        | 81   | 176   | ,85       | 1,49 | 1,07  |
| REFERENCE     | 60        | 41   | 101   | ,53       | ,74  | ,60   |
| Pronominal    | 27        | 13   | 40    | ,24       | ,23  | ,24   |
| Demonstrative | 23        | 23   | 46    | ,20       | ,41  | ,27   |
| Comparative   | 10        | 5    | 15    | ,09       | ,10  | ,09   |
| SUBSTITUTION  | 2         | 4    | 6     | ,02       | ,07  | ,04   |
| Nominal       | 1         | 3    | 4     | ,01       | ,05  | ,02   |
| Verbal        | 1         | 1    | 2     | ,01       | ,02  | ,01   |
| ELLIPSIS      | 2         | 0    | 2     | ,02       | -    | ,01   |
| Nominal       | 2         | 0    | 2     | ,02       | -    | ,01   |
| Verbal        | -         | -    | -     | -         | -    | -     |
| CONJUNCTION   | 26        | 18   | 44    | ,23       | ,35  | ,27   |
| Additive      | 20        | 8    | 28    | ,17       | ,15  | ,17   |
| Adversative   | 3         | 2    | 5     | ,03       | ,04  | ,03   |
| Causal        | 3         | 6    | 9     | ,03       | ,12  | ,06   |
| Temporal      | -         | 2    | 2     | -         | ,04  | ,01   |
| LEXICAL       | 5         | 18   | 23    | ,05       | ,33  | ,15   |
| Repetition    | -         | 8    | 8     | -         | ,14  | ,05   |
| Synonymy      | -         | -    | -     | -         | -    | -     |
| Hyponymy      | -         | 3    | 3     | -         | ,06  | ,02   |
| Collocation   | 5         | 7    | 12    | ,05       | ,13  | ,08   |

the examination texts. This finding points up what we know to be a very serious problem among student writers using a second language as an education medium: severely limited active vocabularies.

TABLE 2: FREQUENCIES OF REFERENCE AND CONJUNCTIVE ERRORS (FUNCTIONAL CATEGORISATION)

|                   | Reference | Conjunctive | Total |
|-------------------|-----------|-------------|-------|
| Extraction        | 51        | 5           | 56    |
| Form              | 4         | 4           | 8     |
| Omission          | 15        | 18          | 33    |
| Replacement       | 3         | 4           | 7     |
| Ambiguity         | 5         | -           | 5     |
| No interpretation | 23        | 13          | 36    |

Table 1 also reveals clearly a very low frequency of substitution and ellipsis errors. This is bound to follow from the very low incidence of these two forms of cohesion in writing as opposed to speaking (Witte and Faigley 1981; Eiler 1983). Lexical errors were more frequent, but reference errors were by far the most frequent, with conjunctive errors next. As indicated earlier, errors in these two categories were analysed further, as shown in Table 2.

The totals in Table 2 reveal three very low frequency (form, replacement and ambiguity) and three very high frequency categories of error (extraction, no interpretation and omission). However, although the most frequent reference errors were extractions, there were only a handful of conjunctive errors of this sort, while the most frequent conjunctive errors were omissions. Perhaps the most disconcerting of all the findings is the fact that "no interpretation" errors were the second most frequent in both the reference and conjunctive categories. Given the assumption that such errors are likely to be the most troublesome to the reader, affecting textual coherence most adversely, the eradication of such errors should be a priority in any pedagogical programme directed at better cohesion in student writing.



## CONCLUSION

The main concern of this study was to set up and apply a framework to the analysis of cohesion errors in student academic writing, and this has resulted in the identification of a number of problem areas. By showing students how the different errors cause different kinds of processing problems for their readers - in other words, by discussing examples and making the categories of the framework explicit to them - we should be able to help them not only to use the cohesion system of English more accurately and effectively, but also to do something that is, especially in writing, at least as important: to develop a sense of audience, a feeling for the reader over one's shoulder.

There are many studies that have shown that cohesion and coherence are by no means the same thing (e.g. Morgan and Sellner 1980: Carrell 1982). There is, though, no doubt that cohesion errors will tend to affect coherence adversely, and that raised awareness of these errors should help teachers to help students to root them out.

We shall, however, have to work hard at it. In **Much ado about nothing**, Dogberry announces that "To be a well-favoured man is the gift of fortune; but to write and read comes by nature." English teachers will be the last to have to be reminded that in this matter, as in others, Dogberry was more than a little wide of the mark.

## BIBLIOGRAPHY

Abbott, G. 1980. "Towards a more rigorous analysis of foreign language errors." In: *International Review of Applied Linguistics* 18 (2): 122-134.

Bamberg, B. 1984. "Assessing coherence: a reanalysis of essays written for the national assessment of educational progress, 1969-1979." *Research in the Teaching of English* 18 (3): 305-319.

- Carrell, P.L. 1982. "Cohesion is not coherence." *TESOL Quarterly* 16 (4): 479-488.
- Chafe, W.L. 1986. "Writing in the perspective of speaking." In: Cooper, C.R. and Greenbaum, S. (eds.) 1986: 12-39.
- Cooper, C.R. and Greenbaum, S. (eds.) 1986. *Studying Writing: Linguistic Approaches*. Beverly Hills: SAGE.
- Freedman, A., Pringle, I., and Yalden, J. 1983. *Learning to Write: First Language/Second Language*. London: Longman.
- Halliday, M.A.K. and Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Hubbard, E.H. 1989. *Reference cohesion, conjunctive cohesion and relational coherence in student academic writing*. Unpublished doctoral thesis, University of South Africa, Pretoria.
- Hunt, K. 1965. *Grammatical Structures Written at Three Grade Levels*. Champaign, Illinois: National Council of Teachers of English.
- Krashen, S.D. 1982. *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Lieber, P.E. 1981. *Cohesion in ESL Students' Expository Writing: A Descriptive Study*. Doctoral thesis, New York University. Ann Arbor: University Microfilms International.
- Morgan, J. and Sellner, M. 1980. "Discourse and linguistic theory." In: Spiro, R.J. et al. (eds.) 1980:165-200.
- Sanford, A.J. and Garrod, S.C. 1981. *Understanding Written Language*. Chichester: John Wiley.
- Spiro, R.J., Bruce, B.C. and Brewer, W.F. (eds.) 1980. *Theoretical Issues in Reading Comprehension*. New York: Erlbaum.

Widdowson, H.G. 1983. "New starts and different kinds of failure." In: Freedman, A. et al (eds.) 1983:34-47.